Beyond maximum likelihood I

Maximum likelihood estimation gives us a single value for the unknown parameters $\boldsymbol{\theta}$, a so-called point estimate.

In many settings in statistical inference we want more than just a point estimation; we want a point estimate *and* some idea of the uncertainty in our point estimate.

For example, when we estimate a single parameter θ, we might want a range of values $[\theta_1, \theta_2]$ which we can reasonably believe that the true value θ lies in.

Beyond maximum likelihood II

Alternatively, we may want to test a hypothesis about θ.

The likelihood function can be used to construct appropriate methods of inference in these settings too.

As with maximum likelihood estimation it can often be shown that they are in some sense optimal.

Interval estimation

Assume, in the one parameter case, that we have a likelihood function $L(\theta; \mathbf{x})$ defined for $\theta \in \Theta$, maximised at its maximum likelihood estimate $\hat{\theta}$.

Then a natural choice of interval estimate is to set some threshold, L_0 say, and to use the values of θ such that $L(\theta; \mathbf{x}) \geq L_0$ as an interval estimate.

Interval Estimation

One natural choice for the threshold is to choose L_0 to be a fixed multiple of the maximum likelihood, say

$$L_0 = e^{-k} L(\hat{\theta}; \mathbf{x})$$

or, equivalently in terms of the log-likelihood,

$$\log L_0 = \ell(\hat{\theta}; \mathbf{x}) - k.$$

Choice of k

Our choice of k involves a trade off between

- an answer that closely matches our data (meaning a smaller interval),
- minimising the risk of missing the true value from the interval (meaning a large interval).

A small k will give a narrow interval but relatively low confidence that the interval contains the true value.

A large k will give a larger interval and higher confidence.

Definition

The k-**unit likelihood region** for parameters $\boldsymbol{\theta}$ based on data $\mathbf{x}$ is the region

$$R_k = \left\{ \boldsymbol{\theta} : \ell(\boldsymbol{\theta}; \mathbf{x}) \geq \ell(\hat{\boldsymbol{\theta}}; \mathbf{x}) - k \right\},$$

or equivalently

$$R_k = \left\{ \boldsymbol{\theta} : L(\boldsymbol{\theta}; \mathbf{x}) \geq e^{-k} L(\hat{\boldsymbol{\theta}}; \mathbf{x}) \right\},$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$ based on $\mathbf{x}$.

k-unit regions

The values of $\boldsymbol{\theta}$ within the k-unit likelihood region are those whose likelihood is at least within a factor e^{-k} of the maximum.

For instance, points in the 1-unit region have likelihoods within a factor $e^{-1} = 0.368$ of the maximum.

The 2-unit region contains points with likelihoods within a factor $e^{-2} = 0.135$ of the maximum.

The 2-unit region is the most commonly used in practice.

Example

Example 37 Interval estimation based on likelihood for normal distributions.

Hypothesis tests

If we are trying to test a null hypothesis $H_0 : \theta = \theta_0$ against a general alternative hypothesis $H_1 : \theta \neq \theta_0$, then we can use a similar idea.

We compare the likelihood of the null hypothesis value, $L(\theta_0; \mathbf{x})$, with the maximum likelihood, $L(\hat{\theta}; \mathbf{x})$ where $\hat{\theta}$ is the maximum likelihood estimate.

If the former is much smaller than the latter, then we reject the null hypothesis; if not we do not.

Precise definition

We can make this precise by saying that we reject the null hypothesis if and only if the ratio

$$\frac{L(\theta_0; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})} \leq e^{-k},$$

where the parameter k is chosen in a similar way to when finding a likelihood region.

Example

Example 38 Hypothesis tests based on likelihood for normal distributions

Going further

This leads into the idea of **likelihood ratio tests**, which you will see more of if you take further courses in statistics.